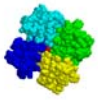


Introduction to Molecular Modeling and Computer Simulation

S. Ravichandran
ABCC, NCI-Frederick,Frederick, MD 21702
E-mail: sravi@ncifcrf.gov
Tel: X1991
http://nciiris.ncifcrf.gov/~ravichas/MM/MM.htm

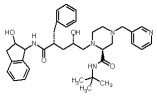
Overview of the class

- What is Molecular Modeling?
  - Basic assumptions, graphics etc.
- Properties: Hydrophobicity, ASA, electrostatics etc.
- Techniques: MM, MD, QM
  - QSAR, Protein-ligand docking
- Hands on exercise:
  - Small Molecule building, Energy Minimization
  - Aligning protein structures using Homology
  - Pharmacaphore concepts
  - Protein-Ligand Docking

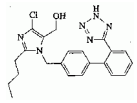
Capabilities/Limitations Drug Design Approach

- *"It is impossible to expect a molecular structure to appear on the screen of a workstation simply by asking the computer for a novel molecule that will cure cancer, is inexpensive to manufacture, and has no side-effects. Rarely does computational chemistry lead directly to drug"*
  - » Donald B. Boyd, Indiana University, Drug Design

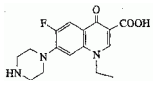
Indinavir: Antiviral Drug/HIV/MM/MD/ Crystallography



Losartan: Antihypertensive agent/Structure Activity Study



Norfloxacin:Antibiotic /QSAR

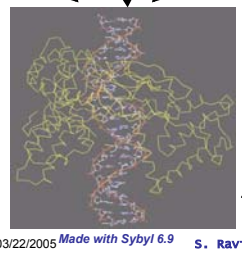


Molecular Modeling

Molecular Modeling is the science of creating/studying molecular structure and function through models and simulation

Sketching **Hom. Mod.** **Experiments**

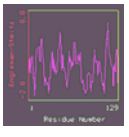
GRAPHICS



Rotation/Translation

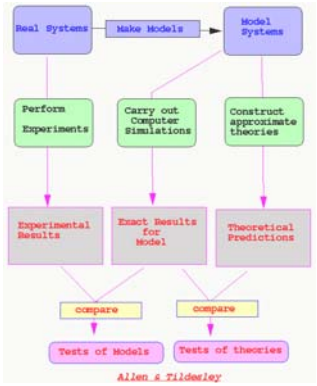
Calculation/Analysis

Simulation, Analysis



Made with InsightII

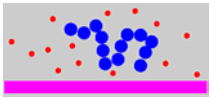
Made with Sybyl 6.9



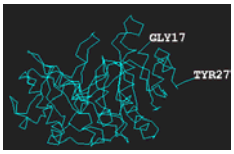
Molecular Modeling

Model to study protein-surface adsorption

Alpha chain model



Why Molecular Modeling?



1WSA shown using Sybyl 6.9

26 Missing residues-a flexible loop

Loop is found to be important for the functioning of this molecule.

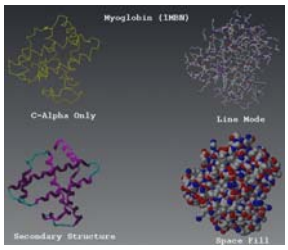
Advantage: 3-D structure availability
Dis-advantage: incomplete structure

How do we get the complete structure?

Molecular Modeling (bioinformatics, Homology modeling…..)

Applications of Molecular Modeling

- Visualization
  - Free: Spdbv, Cn3D, Rasmol, VMD and many more (**Exercise: Spdbv, Cn3d**)
  - $$$$: Tripos, Accelrys and many more



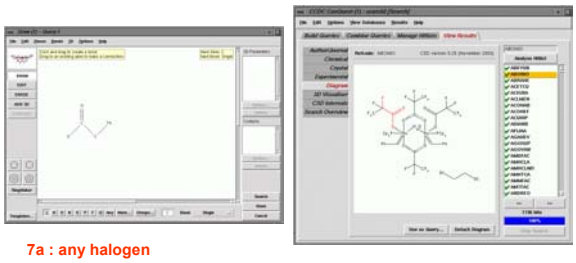
Cambridge Structural Database

- X-ray and neutron diffraction analysis of carbon-containing molecules (up to 1000 atoms including H)
  - Organics, Organometallics, Metal Complexes
  - Peptides up to 24 residues
  - mono-, di- and tri-nucleotides
- Different Search Options:
  - Basic substructure, Substructure with constraints, 3D substructure, non-bonded interactions, Pharmacophore, Cell parameter, Journal Reference

CSD: Substructure Search

Query

Results



7a : any halogen

3D-Structural Database http://www.rcsb.org

- NMR
  - Dynamic
    - **Multiple Models (Each conformation is a model)**
  - Aqueous environment
  - Limitations
    - Size of molecule
      - < 30kD
- Example
  - 1DV0, 1UBA

- X-ray
  - Static
    - Only one model
  - Crystal
  - Limitations
    - Not limited by size
- Examples
  - 7LYZ

Structural Databases

- These crystallographic databases gives information w.r.t a crystal environment
  - Proteins NMR studies have proved that the structure in the crystal phase and solution phase are almost same but for small molecules this may not be the case
  - These databases do not cover the whole spectrum because some of the molecules cannot be crystallized

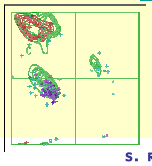
No Experimental Structure & Homology Modeling

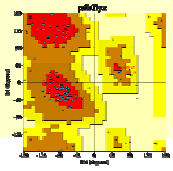
- No 3D structure but has homologous PDB entries
  - Can exploit homology to model the unknown protein
    - Accelrys (Modeller), Swiss-Model, Tripos (Match-maker,)
- No 3D structure but do not have any homologous PDB entries
  - Threading, Reverse Folding
    - Tripos (GenFold)

Quality (model) check!

- Procheck: Stereo-chemical quality of the protein and residue by residue analysis in figures

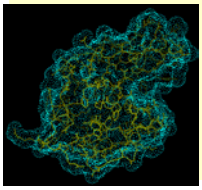
  http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
- PDBREPORT: http://www.cmbi.kun.nl/gv/pdbreport

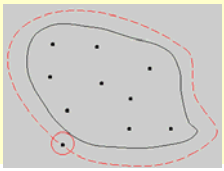




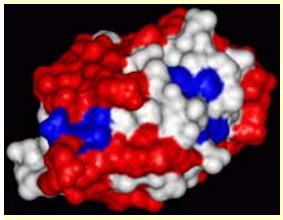
Important properties

- Solvent-Accessibility (SA)
  - SA help us to know what groups are on the surface-solvent exposed
    - Can give hints on the possible interaction with ligands etc.

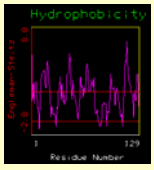




Molecular Surface



Hydrophobic residues on the surface of the protein for some reason! Protein-Protein Interactions

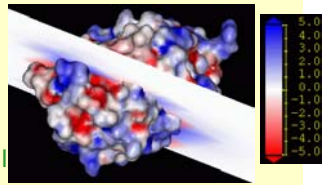


Lysozyme, Hydrophilic red, hydrophobic blue, InsightII, subsets are created using Engleman-Steitz algorithm

Electrostatics

- Delphi: software to calculate electrostatic properties-Protein-Ligand interactions
  - Calculate electrostatic potential
  - Effects of site-directed mutagen-esis
  - Electrostatic con-tribution to the solvation energy
  - Interface to InsightII

Picture made with InsightII



Computer Simulation

Molecular Mechanics (MM)

- What is Molecular Mechanics?
  - MM is a energy refinement procedure. Refinement process Is usually called Minimization or Energy Minimization.
  - Assumption: Energy minimized structure is closer to the stable geometry and probably closer to experimental structure.
- Where Energy Minimization is usually employed?
  - Molecule Building, Homology modeling, Conformational Search, PDB file refinement

MM

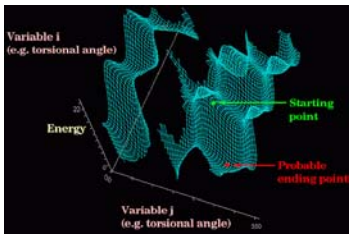
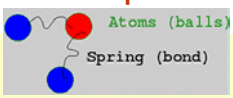


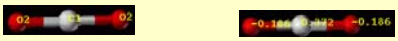
Image from NIH Molecular Modeling pages

Basic assumptions of MM



- Electrons and nuclei are lumped together
- Molecules are assumed to be balls (point masses) and connected to others by bonds (springs)
- Total energy of the system is an important property and it is usually computed as a sum of independent energy terms.
  - Electrostatic energy term: $E\_ele = (q_i q_j)/(4\pi\varepsilon_0 r^2)$

MM



- Each atom/bond in a molecule or amino acid is identified by
  - 1) Atom type 2) Residue type 3) hybridization type 4) Bonding info. 5) Charges 6) Coordinates

Carbon-di-oxide

- Atom types C1, O2; Hybridization: sp, sp2; Bonding info: C1 is bonded to two O2 atoms; Coordinates: x,y and z; Charges: (C) 0.372, (O) -0.186

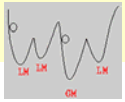
ForceField

- ForceField is an analytical Functional form for the independent energy terms and parameters

Bond length = $K(b-b_0)^2$
Simple Functional Form

$$E_{pot} = \Sigma 1/2K_b(b-b_0)^2 + \Sigma 1/2K_\theta(\theta-\theta_0)^2 + \Sigma 1/2K_\phi(1+\cos N\phi)^2 + \Sigma 1/2K_\chi(\chi-\chi_0)^2 \Sigma((B/r)^{12} - (A/r)^6) + \Sigma(qq/r)$$

- Different Flavours of Energy Minimization:
  - Steepest Descent (SD) and Conjugate-Gradient (CD)
    - SD is used to relieve overlaps and so good at start
    - CD is slow but can lead to structures with low energies. Do not get trapped in local minima like SD!
    - Simulated Annealing



MM

- Limitations: Not possible to reach Global minimium.
  - Two alternative methods:
    MD or stepwise rotation of bonds

Example

Twist boat cyclohexane
11.917 kcal/mol

Molecular Dynamics →

Chair form
6.558 kcal/mol

Cyclohexane remains in twist boat form in Molecular Mechanics

Molecular Dynamics (MD)

- Time dependent behavior of the molecular system
  - Local vibrations, conformational change of proteins and nucleic acids
- MD is based on classical Newton's motion
  - Equation of motion: F = m x a
- A typical MD run consist of the following steps
  - Set Initial configuration/Velocity; Heating, Equilibration, Production, Saving configurations
- Applications: Dynamical Properties, MD can take information from NMR to perform a restrained MD.

Gromacs, Amber, Charmm, VMD, NEMD

Quantum Mechanics (QM)

- Need for QM
  - MM and MD do not consider electrons explicitly (Born-Oppenheimer approximation)
    - When a drug molecule interact with a receptor. Primary interactions occur between the electron clouds. ELECTRONIC influence cannot be ignored always.
  - MM and MD cannot answer questions related to
    - Bond-forming or bond-breaking
    - Molecules not in ground state

QM

- Basics: $H\Psi = E\Psi$ Shrodinger's Equation E=Energy,Ψ = Wave Function.
  - Solve S.E to get the Energy and Wave Function, which inturn can be used to extract electronic properties (electron density etc.)
  - *ab-initio*, semi-empirical (AM1, etc.)
- QM can be used in conformational search and energy minimization
- Flavors: MOPAC, GAMESS etc.
- Applications: Minimization for small molecules, for conjugated systems, Descriptors for QSAR, Partial charges, transitition state geometries & energies

Advanced Techniques

- MC (Monte-Carlo), Brownian Dynamics, QM/MD

Advanced Topics

- Protein-ligand Docking
  - AutoDock
- QSAR

Protein-Flexible ligand Docking

Docking

- Fitting a small molecule (drug molecule) into a protein
- Docking two proteins together
- Things you need:
  - Protein PDB X-ray/NMR
    - Homolgy Model Databases
  - Small Molecules CSD X-ray
    - Sketching
  - Software: Dock, AutoDock, Ludi, FlexX etc
  - Disk-space, CPU .....

Docking

- Attempt to identify the binding mode of L-R complex by searching conf/orient. space for a geometry with favourable binding energy
- Questions to ask before you begin
  - How good is the protein structure?
    - Resolution (~2 Å), R-factor (below 20%), free R-factor below 30%, B-factor (active site), experimental info. 40 $Å^2$ indicate problems, Crystal packing forces, Symmetry related copies of protein may influence protein conformation

Docking

- NMR 1model/multi-models, which one to keep?
  - High resolution structure should have an order of 20 (distance or dihedral) restraints/residue
  - RMSD variation between ensemble of structures
    - < 1 Å backbone atoms
    - < 1.5 Å all atoms
- Are there bound ions/water mols/ligands?
  - Remove them/keep them ??
- Is the crystal structure appropriate for your modeling experiment ?
  - Inhibitor for kinases-which form active/inactive?

Docking

Basics:

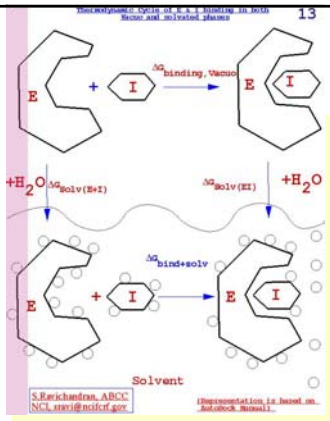
E + I $\rightleftharpoons$ E--I (loosely bound complex)

Why is it difficult?

2 rigid molecules-6 degrees of freedom

14 rotatable bonds

1028 variations

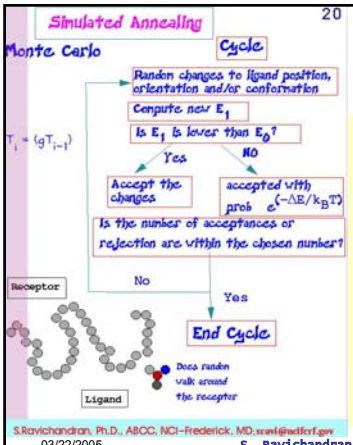


Docking

Thermodynamic Cycle approach

Popular Methods of Docking

1. Simulated Annealing
2. Genetic Algorithms



Docking

Annealing: A process in which the T of the system is reduced (slowly) until the material crystallizes in a singe crystal (correspond to global FE)

SA can do both global and local search

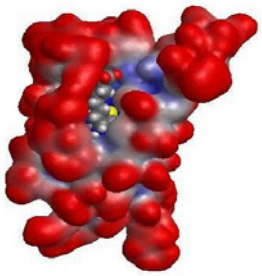
Global: High T

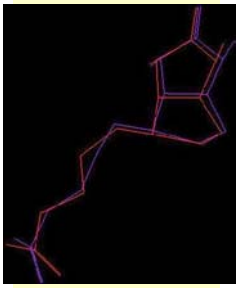
Local : Low T

Test Case Using AutoDock

Streptavidin/Biotin complex (1stp)

P.C. Weber, D.H. Ohlendorf, J.J. Mendolowski, and F.R. Salemme (1992)





QSAR

- In the process of searching for lead compounds
  - ∞ number of possible analogues can be made
    - Substituents on aromatic, functional groups etc.

QSAR

- What is QSAR

  Quantitative Structure Activity Relationships

Addresses two questions:

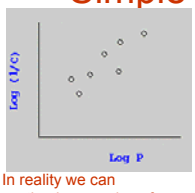
- What feature of a molecule affect its activity?
- What can be modified to enhance properties?

Quantitative in that a mathematical model is used to account for the observed activity

Applications of QSAR

- Drug Design
  - Predictions for new experiments
  - Correlate different kinds of biological activity
  - Elucidate the mechanism of new drugs

Simple Linear Regression

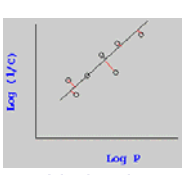


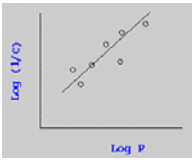
In reality we can synthesize a series of compounds which can affect only Log (p)

Activity expressed as 1/C
C = concentration of a drug molecule required to produce expected level of biological activity

Idea is to see whether there is any relationship between x & y

Why 1/C? Most active achieve defined activity at a lower C.





$Y = k_1x + k_2$

k_1 and k_2 are constants

Physicochemical Properties

- Hydrophobicity, Electronic & Steric
- Quantitative description of hydrophobicity is comparatively easy
  - Partition coefficients (log P) or π (hydrophobic)

$$P = \frac{\text{Concentration of drug in octanol}}{\text{Concentration of drug in aq. soln.}}$$

Hydrophobic (High P); Hydrophilic (Low P)

Binding of drugs to Serum Albumin

log(1/c) = 0.75 log P + 2.3 (based on 40 Compounds)

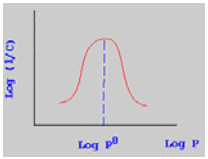
Binding is determined mainly by hydrophobicity.

Is this true for all P values?

True for small ranges of P (1-4). What happens at high P values?

log (p)

Parabolic Curve



Does the log (p) increases infinitively?
Reaches optimal value and decreases.
Can we describe this by an equation?
Parabolic Curve
What happens at higher P values?

Why do we see such as behavior?

Are there drugs which depend only on log P?

Yes, Anaesthetics which enter cell-membranes and affect the CNS activity- No drug-receptor interactions

General Equation;

$Y = (-)\, k_1(\log P)^2 + k_2 \log P + k_3$

$Y = -0.22\,(\log P)^2 + 1.04 \log P + 2.16$

Gl. Eq. for range of anaesthetic ethers

Can we extent it for any anaesthetics?

log (P)

- It has been shown that any compound with a log P close to 2 can efficiently enter CNS and act as anaesthetic
- What would you do if you want your drug molecule to get stuck in the CNS?

Substituent Hydrophobicity constant (π)

π_{ali} Aliphatic Compunds
π_{Ar} Aliphatic Compunds

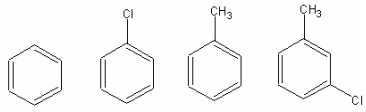
- Log (p) values represent the hydrophobicity for the whole compound. To calculate log(p) we need to synthesize the molecules.
  - Is there anyway to quantify hydrophobic effects of functional groups?
- $\pi = \log P_X - \log P_H$

X	CH_3	t-Bu	OH	OMe	CF_3	Cl	Br	F
π_{ali}	0.5	1.68	-1.16	0.47	1.07	0.39	0.60	-0.17
π_{Ar}	0.52	1.68	-0.67	-0.02	1.16	0.71	0.86	0.14

How does this solve the problem?

Still the lead compound log P has to be determined experimentally. Once it is known then the analogues can be calculated using π

Substituent Hydrophobicity constant (π)



Log(p) Experimental value 3.28

Values are also available specifically for meta substituents

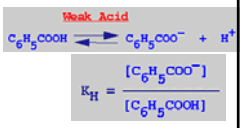
2.13	0.71	0.56	3.40
log(p)	π_{Cl}	π_{Me}	log(p)

Hugo Kumbinyi" QSAR Parameters"

Quantum Mechanical Descriptors: HOMO (Ionization Potential), LUMO (e- affinity) etc.

Electronic Effects

- Hammett substitution constant σ
- It is a measure of electron-withdrawing or electron donating ability of a substituent
- For substituent benzoic acids
  - Hammett S.C: $\sigma_X = K_X/K_H$



Molar Refractivity (MR): Bulkness of the drug molecule

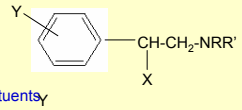
MR = [(n2 -1)/(n2+2)] (MW/ρ)

n refractive index, MW Molecular Weight, ρ Density

Hansch Equation

- Relation of biological activity to more than one parameter
  - QSAR Equation: log(1/C) = 1.22 π - 1.59σ + 7.89

Electronic factor



Adrenergic blocking activity of β-halo-arylamines

Log(1/C) increases if the substituents has + π value and –σ value (Hydrophobic & Electron donating)

Statistical Analysis

- R-squared (r2)
  - Is the correlation coefficient
  - Goodness of a fit

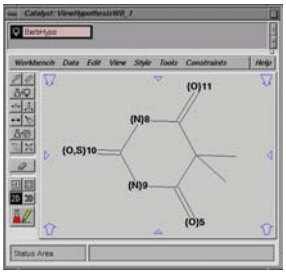
$$r^2 = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - Y_{mean})^2}$$

= 1 perfect fit

= >= 0.95 good model

= <= 0.70 Poor model

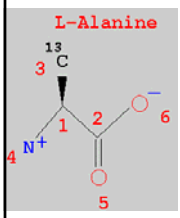
Substructure Search



Substructure for generic barbiturate

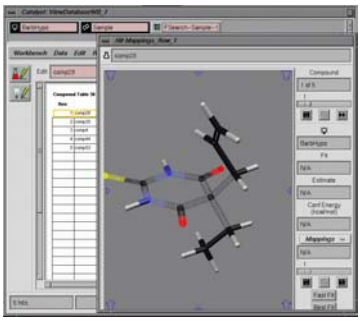
Catalyst 4.9 Accelrys Inc.

MDL 2D file format (CTFILE)



Connection Table (Ctab)

Substructure Search



Hits from the Fast Flexible Search Database/Spreadsheet Search on a sample database

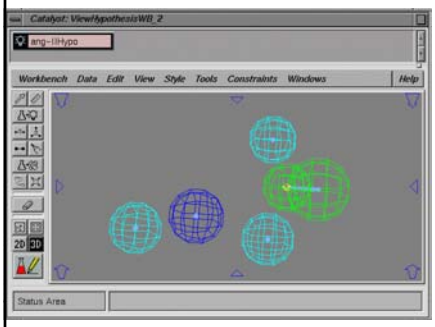
Options to save the hits

Fit the hits to the hypothesis

.I……

……..

3D-Search



Cyan Hydrophobes

Blue Neg Ionizable

Green HB acceptor

Once you have identified a set of compounds that exhibited activity for the same assay. Use them to generate a 3D hypothesis

Hands-on Exercise

- Instructions in the web-link
  - http://nciiris.ncifcrf.gov/~ravichas/MM/MM.htm

Selected Reference Books (Molecular Modeling)

- Molecular Modeling and Simulation, T. Schlick (2002)
- Molecular Modeling: Principles and Applications A.R. Leach (2001)
- Computer Simulation of liquids, M.P. Allen and D.J. Tildesley (1989)
- Bioinformatics: A practical Guide to the analysis of Genes and Proteins, Edited by A.D. Baxevanis and B.F.F. Quellette (2001)
- Bioinformatics Basics, H. H. Rashidi and L.K. Buehler (2000)

Selected Reference Books (Molecular Modeling)

- Developing Bioinformatics Computer Skills, C.Gibas and P. Jambeck (2001)
- Introduction to Bioinformatics: Atwood and Parry-Smith (1999)
- Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Andreas D. Baxevanis, B.F. Oullette (2001)
- Introduction to Bioinformatics, Arthur M.Lesk (2002)
- Discovering Genomics, Proteomics & Bioinformatics, A. M. Campbell and L. J. Heyer (2003)

Selected Reference Books (QSAR)

- **Molecular Modelling: Principles and Applications**, 2nd Edition, Andrew R. Leach
- **Encyclopedia of Computational Chemistry, 5 Volume Set** by Paul Von R. Schleyer (Editor), Paul Von Rague Schleyer
- **An introduction of QSAR Methodology**, Allen B. Richon and Stanley S. Young, Network Science (1997)